

明 細 書

メタデータ作成装置、その作成方法および検索装置

技術分野

本発明は、制作された映像、音声等のコンテンツに関するメタデータを作成するためのメタデータ作成装置、及びメタデータ作成方法に関する。また、作成されたメタデータによりコンテンツを検索する検索装置に関する。

背景技術

近年、制作された映像、音声コンテンツに対して、それらのコンテンツに関するメタデータの付与が行われている。

しかしながら、従来のメタデータ付与の作業は、制作された映像、音声コンテンツのシナリオあるいはナレーション原稿をもとに、制作された映像、音声コンテンツを再生しながらメタデータとすべき情報を確認し、手作業でコンピュータ入力することにより作成する方法が一般的であった。従って、メタデータの作成には、相当な労力が必要であった。

特開平09-130736号公報には、カメラ撮影時に音声認識を行いタグ付けをするシステムが記載されている。しかしながら、このシステムは、撮影と同時に使用されるものであり、制作されたコンテンツに対してメタデータを付与するために適用できるものではなかった。

発明の開示

本発明は、上記課題を解決するものであって、制作されたコンテンツに対して、音声入力により容易にメタデータを作成可能とするメタデータ

タ作成装置、及びメタデータ作成方法を提供することを目的とする。

また、そのようにして作成されたメタデータを用いて、コンテンツの内容を容易に検索可能な検索装置を提供することを目的とする。

本発明のメタデータ作成装置は、コンテンツを再生して出力するコンテンツ再生部と、音声入力部と、前記音声入力部から入力された音声信号を認識する音声認識部と、前記音声認識部により認識された情報をメタデータに変換するメタデータ生成部と、前記コンテンツ内の各部位を識別するための識別情報を、前記コンテンツ再生部から供給される再生されたコンテンツから取得して前記メタデータに付与する識別情報付与部とを備え、生成された前記メタデータを前記コンテンツ内の各部位と関連付けるように構成される。

本発明のメタデータ作成方法においては、コンテンツに関連した情報を音声入力し、入力された音声信号を音声認識装置により音声認識し、前記音声認識された情報をメタデータに変換し、前記コンテンツ内の各部位を識別するための前記コンテンツに付与された識別情報を前記メタデータに付与して、生成された前記メタデータを前記コンテンツ内の各部位と関連付ける。

本発明のメタデータ検索装置は、コンテンツを再生して出力するコンテンツデータベースと、入力されるキーワードの音声信号を再生された前記コンテンツの同期信号に同期したクロックでデータ化する音声入力部と、前記音声入力部でデータ化された音声信号データからキーワードを認識する音声認識部と、前記音声認識部から出力されたキーワードを前記コンテンツに含まれる画像信号の時間位置を示すタイムコードと結合させてメタデータファイルを作成するファイル処理部と、コンテンツファイルの記録位置と前記メタデータファイルの関係を管理する制御ファイルを発生するコンテンツ情報ファイル処理部と、前記コンテンツフ

5 アイル、前記メタデータファイル、および前記制御ファイルを記録する記録部と、入力される検索用キーワードが含まれた前記メタデータファイルを特定し、前記制御ファイルを参照して、前記コンテンツファイルの前記キーワードに対応する記録位置を抽出する検索部とを備える。前記コンテンツファイルの記録位置は、前記記録部における記録位置である。

図面の簡単な説明

10 図 1 は、本発明の実施の形態 1 に係るメタデータ作成装置の構成を示すブロック図である。

図 2 は、本発明の実施形態 1 に係るタイムコード付与メタデータの一例を示す図である。

図 3 は、本発明の実施形態 2 に係るメタデータ作成装置の構成を示すブロック図である。

15 図 4 は、同装置における静止画コンテンツ・メタデータ表示部の一例を示す図である。

図 5 は、本発明の実施形態 2 に係るメタデータ作成装置の他の構成を示すブロック図である。

20 図 6 は、本発明の実施の形態 3 に係るメタデータ作成装置の構成を示すブロック図である。

図 7 は、同実施の形態の装置における辞書 D B の一例を示す構成図である。

図 8 は、同実施の形態の装置に適用されるコンテンツのシナリオの一例であるレシピを示す図である。

25 図 9 は、同実施の形態の装置により作成されるメタデータファイルの一例を示す TEXT 形式のデータ図である。

図10は、本発明の実施の形態4に係るメタデータ作成装置の構成を示すブロック図である。

図11は、同実施の形態の装置により作成される情報ファイルの一例を示す構成図である。

5 図12は、本発明の実施の形態5に係るメタデータ検索装置の構成を示すブロック図である。

図13は、本発明の実施形態6に係るメタデータ作成装置の構成を示すブロック図である。

10 発明を実施するための最良の形態

本発明のメタデータ作成装置によれば、コンテンツに関連したメタデータの作成あるいはタグ付けを行う際に、音声認識を用いて音声入力によりメタデータあるいはタグを作成するとともに、メタデータあるいはタグとコンテンツの時間あるいはシーン等との関連付けを行う。それにより、従来キーボードで入力して作成していたメタデータを、音声入力により自動的に作成することが可能となる。なお、メタデータとはタグの集合体を意味し、この本発明においてメタデータと称する場合は、タグ自身を意味する場合も含む。また、コンテンツとは、制作された映像、音声コンテンツ、静止画コンテンツ、データベース化された映像、音声コンテンツ等、一般にコンテンツと称されるあらゆるものを含む意味で用いられる。

本発明のメタデータ作成装置は、好ましくは、コンテンツに関連した辞書を更に備え、前記音声入力部から入力された音声信号を前記音声認識部により認識する際に、前記辞書に関連付けて認識するように構成される。この構成によれば、制作されたコンテンツのシナリオ等から事前に抽出したキーワードを音声信号として入力し、また、シナリオに基づ

いて辞書分野の設定、及び、キーワードの優先順位つけを行うことにより、効率よく、正確に音声認識手段からメタデータを発生することができる。

さらに、前記音声信号を前記音声認識部により前記辞書に関連付けて
5 単語単位で認識する構成としてもよい。また好ましくは、キーボードを含む情報処理部を更に備え、前記キーボードからの入力により前記情報処理部を介して前記メタデータを修正可能とする。前記識別情報として、前記コンテンツに付与されているタイムコード情報を用いることができる。あるいは、前記識別情報として、前記コンテンツに付与されている
10 コンテンツの番地、番号あるいはフレームナンバを用いてもよい。また、前記コンテンツが静止画コンテンツであり、前記識別情報として、前記静止画コンテンツの各々の番地を用いることができる。

本発明の一適用例として、次のようなメタデータ作成装置を構成することができる。すなわち、前記コンテンツ再生部はコンテンツデータベースにより構成され、前記音声入力部は、入力されるキーワードの音声信号を、前記コンテンツデータベースから供給される同期信号に同期したクロックでデータ化して前記音声認識部に供給する。前記音声認識部は、前記音声入力部でデータ化された音声信号データからキーワードを認識するように構成される。前記メタデータ生成部は、前記識別情報として前記コンテンツに含まれる画像信号の時間位置を示すタイムコードを用い、前記音声認識部から出力されたキーワードを前記タイムコードと結合してメタデータファイルを作成するファイル処理部として構成される。

この構成によれば、数秒単位の間隔であっても効率よくメタデータを付与できる。従って、従来のキー入力では困難であった、短い時間間隔でのメタデータ作成が可能となる。

この構成において好ましくは、前記メタデータファイルとともに、前記コンテンツデータベースから供給されるコンテンツをコンテンツファイルとして記録する記録部を更に備える。また好ましくは、前記コンテンツファイルを記録すべき記録位置と前記メタデータファイルの関係を
5 管理する制御ファイルを発生するコンテンツ情報ファイル処理部を更に備え、前記記録部に、前記コンテンツファイルおよび前記メタデータファイルとともに、前記制御ファイルを記録するように構成する。また好ましくは、辞書データベースを更に備え、前記音声認識部は、複数のジャンル別辞書から、前記コンテンツに適合したジャンルの辞書を選択す
10 ることが可能である構成とする。さらに好ましくは、コンテンツに関連したキーワードを前記音声認識部に供給可能であり、前記音声認識部は、前記キーワードを優先認識するように構成する。

本発明のメタデータ作成方法において好ましくは、前記コンテンツを再生しモニタに表示しながら、コンテンツに関連した情報を音声入力する。また好ましくは、前記コンテンツに関連した辞書を用い、前記入力された音声信号を前記音声認識装置により前記辞書に関連付けて認識する。また好ましくは、前記識別情報として、前記コンテンツに付与されているタイムコード情報を用いる。また、前記コンテンツとして静止画コンテンツを用い、前記識別情報として、前記静止画コンテンツの各々
20 の番地を用いることができる。

本発明のメタデータ検索装置によれば、コンテンツの記録位置を示す制御ファイルとメタデータ、及び、タイムコード等を示すメタデータファイルを使用することによって、メタデータに基づきコンテンツの所望の箇所を高速で検索することができる。

25 本発明のメタデータ検索装置において好ましくは、前記コンテンツ情報ファイル処理部から出力される制御ファイルは、前記コンテンツの記

録時間に合わせた前記記録部における前記コンテンツ記録位置を明記するテーブルとし、前記タイムコードから前記コンテンツの記録位置を検索できるようにする。

また好ましくは、辞書データベースと、コンテンツに関連したキーワードを前記音声認識部に供給するキーワード供給部とを更に備え、前記音声認識部は、複数のジャンル別辞書から、前記コンテンツに適合したジャンルの辞書を選択することが可能であるとともに、前記キーワードを優先認識するように構成する。

また好ましくは、辞書データベースを更に備え、前記音声認識部は、複数のジャンル別辞書から、前記コンテンツに適合したジャンルの辞書を選択することが可能であるとともに、前記検索部は、前記音声認識部で用いた共通辞書から選定されたキーワードにより検索を行うように構成される。

以下、本発明の実施の形態について、図面を参照して具体的に説明する。

(実施の形態 1)

図1は、本発明の実施の形態1におけるメタデータ作成装置の構成を示すブロック図である。コンテンツ再生部1は、メタデータの作成に際して、制作されたコンテンツを確認するための要素である。コンテンツ再生部1の出力は、映像モニタ2、音声モニタ3およびタイムコード付与部7に供給される。メタデータ作成のための音声入力部として、マイク4が設けられる。マイク4から入力された音声は、音声認識部5に入力される。音声認識部5には、音声認識用の辞書8が接続され、そのデータを参照することが可能となっている。音声認識部5の認識出力はメタデータ生成部6に供給され、作成されたメタデータは、タイムコード付与部7に供給され、タイムコード付与部7から外部に出力可能である。

コンテンツ再生部1としては、例えば、VTR、ハードディスク装置、光学ディスク装置等の映像・音声信号再生装置、半導体メモリなどのメモリ手段を記録媒体とする映像・音声信号再生装置、あるいは伝送、放送により供給された映像・音声信号を再生する映像・音声再生装置など
5 が用いられる。

コンテンツ再生部1の映像信号出力端子1aからは、再生された映像信号が映像モニタ2に供給される。音声信号出力端子1bからは、再生された音声信号が音声モニタ3に供給される。タイムコード出力端子1cからは、再生されたタイムコードがタイムコード付与部7に供給される。
10 なお、映像モニタ2および音声モニタ3は、メタデータ作成装置の要素として必須ではなく、必要に応じて接続して使用可能であればよい。

メタデータを作成する際の作業者は、映像モニタ2と音声モニタ3のいずれかあるいは両方を確認しながら、場合によればシナリオまたはナレーション原稿なども参照しながら、入力すべきメタデータを発声マイク4により入力する。マイク4から出力される音声信号は、音声認識部5に供給される。また、必要に応じて、音声認識用の辞書8のデータが音声認識部5により参照される。音声認識部5で認識された音声データは、メタデータ生成部6に供給され、メタデータに変換される。
15

このようにして生成されたメタデータは、コンテンツの各部分の時間あるいはシーンとの関係を対応付ける情報を付与するため、タイムコード付与部7において、再生されたコンテンツから取得され、コンテンツ再生部1から供給されるタイムコード情報が付与される。
20

以上の動作をより具体的に説明するため、コンテンツが料理説明の場合の一場面を例として想定する。この場合オペレータが、映像モニタ2の表示画面を確認しながら、マイク4を通じて"塩 1さじ"と発声すると、音声認識部5では、辞書8を参照しながら、"塩"、"1さじ"
25

と認識され、メタデータ生成部 6 にて各々 "塩"、"1さじ" というタグに変換される。なお、音声認識部 5 の構成は特に限定されるものではなく、一般的に用いられている種々の音声認識手段を用いて音声認識し、"しお"、"ひとさじ" のようなデータを認識すればよい。なお、一般 5 には、メタデータとは、このようなタグの集合体を意味するものである。

音声認識の結果、図 2 に示すように、音声認識部 5 からメタデータ 9 a が output され、タイムコード付与部 7 に供給される。

タイムコード付与部 7 では、コンテンツ再生部 1 から供給されるタイムコード信号 9 b をもとに、タイムコードが付与されたタイムコード付与メタデータ 10 からなるパケットデータが生成される。生成されたメタデータは、そのまま出力されても良く、またハードディスク等の記録媒体に蓄積しても良い。

なお、この例ではパケット形式のメタデータを生成する場合を示したが、これに限定されるものではない。

15 (実施の形態 2)

図 3 は、本発明の実施の形態 2 におけるメタデータ作成装置の構成を示すブロック図である。本実施の形態は、静止画コンテンツをメタデータ作成の対象とする例である。静止画コンテンツを識別するために、動画の場合のタイムコードに相当するコンテンツの番地を用いて、生成されたメタデータと静止画コンテンツとを関連づける構成を有する。

図 3において、カメラ 1 1 は静止画コンテンツ制作のための要素である。カメラ 1 1 の出力は、静止画コンテンツ記録部 1 2 により、番地情報が付与されて記録される。ここで記録された静止画コンテンツおよび番地情報は、メタデータ作成のために、静止画コンテンツ・メタデータ記録部 1 3 に供給される。番地情報はさらに、メタデータ番地付与部 1 9 にも供給される。

マイク 1 6 は、静止画に関連する情報を音声入力するために用いられ、その出力は音声認識部 1 7 に入力される。音声認識部 1 7 には、音声認識用の辞書 2 0 が接続され、そのデータを参照することが可能になって いる。音声認識部 1 7 の認識出力はメタデータ生成部 1 8 に供給され、
5 作成されたメタデータは、メタデータ番地付与部 1 9 に供給される。

静止画コンテンツ・メタデータ記録部 1 3 に記録された静止画コンテ ンツおよびメタデータは、静止画コンテンツ・メタデータ再生部 1 4 に より再生して、静止画コンテンツ・メタデータ表示部 1 5 に表示される。

以上の構成のメタデータ作成装置の動作について、さらに詳細に説明
10 する。

カメラ 1 1 で撮影された静止画コンテンツは、静止画コンテンツ記録 部 1 2 で、記録媒体（図示せず）に記録されるとともに番地情報が付与 され、番地情報も記録媒体に記録される。記録媒体は、一般的には半導 体メモリで構成されるが、半導体メモリに限定されず、たとえば、磁気 メモリ、光記録媒体、光磁気記録媒体など、種々の記録媒体を用いるこ とが可能である。記録された静止画コンテンツは出力端子 1 2 a、入力 端子 1 3 a を介して、同様に、番地情報は出力端子 1 2 b、入力端子 1 3 b を介して、静止画コンテンツ・メタデータ記録部 1 3 に供給される。 番地情報はさらに、出力端子 1 2 b、入力端子 1 9 b を介してメタデ 20 タ番地付与部 1 9 に供給される。

一方、カメラ 1 1 で撮影された静止画に関連する情報が、マイク 1 6 を介して音声認識部 1 7 に入力される。静止画に関連する情報は、例え ば、タイトル、撮影の日時、撮影者、撮影場所（どこ）、被撮影者（だれ）、 被撮影物（なに）などである。また、必要に応じて、音声認識部 1 7 で 25 は音声認識用の辞書 2 0 のデータも供給される。

音声認識部 1 7 で認識された音声データは、メタデータ生成部 1 8 に

供給され、メタデータあるいはタグに変換される。なお、一般には、メタデータとは、コンテンツに関する情報である、タイトル、撮影の日時、撮影者、撮影場所（どこ）、被撮影者（だれ）、被撮影物（なに）等のタグの集合体を意味する。このようにして生成されたメタデータあるいはタグは、静止画コンテンツの内容あるいはシーンとの関係を対応付ける情報を付与するため、メタデータ番地付与部19に供給される。メタデータ番地付与部19において、静止画コンテンツ記録部12から供給された番地情報が、メタデータに付与される。この様にして番地情報の付与された番地付与メタデータは、出力端子19c、入力端子13cを介して、静止画コンテンツ・メタデータ記録部13に供給される。静止画コンテンツ・メタデータ記録部13では、同一の番地の静止画コンテンツと同一の番地のメタデータとが関連付けて記録される。

番地付与メタデータについてより具体的に説明するため、静止画コンテンツ・メタデータ記録部13により記録された静止画コンテンツおよびメタデータを、静止画コンテンツ・メタデータ再生部14により再生し、静止画コンテンツ・メタデータ表示部15により表示した結果の一例を、図4に示す。

図4における、静止画コンテンツ・メタデータ表示部15の画面は、一例であるが、静止画コンテンツの表示部21、番地の表示部22およびメタデータの表示領域23から構成される。メタデータの表示領域23は、たとえば1) タイトルの記述部23a、2) 日時の記述部23b、3) 撮影者の記述部23c、4) 撮影場所の記述部23dなどで構成される。これらのメタデータが、上述の音声認識部17により認識された音声データから生成される。

上述の動作は、メタデータの生成を、静止画コンテンツの撮影の事前、撮影とほぼ同時期あるいは、撮影の直後など、必ずしも、撮影された静

止画コンテンツの確認を必要としない場合に関するものである。

次に、静止画コンテンツの後付としてメタデータの生成を行う場合に、静止画コンテンツを再生し、モニタした静止画コンテンツに対して、メタデータを生成する場合について、図5を参照して説明する。なお、図5と同様の要素については、同一の番号を付してその機能等の説明を省略する。この場合は、静止画コンテンツ記録部12と静止画コンテンツ・メタデータ記録部13の間に、静止画コンテンツ・番地再生部24が設けられる。更に、静止画コンテンツ・番地再生部24の出力が供給されるモニタ25も設けられる。

カメラ11で撮影され、静止画コンテンツ記録部12に供給された静止画コンテンツは、記録媒体（図示せず）に記録されるとともに番地が付与され、番地も記録媒体に記録される。このような記録媒体が、静止画コンテンツ・番地再生部24に供給される。従って、このように、制作済の静止画コンテンツを再生し、モニタした静止画コンテンツに対して、メタデータを生成するために用いられるメタデータ作成装置では、カメラ11および静止画コンテンツ記録部12は必須の要素ではない。

静止画コンテンツ・番地再生部24で再生された静止画コンテンツは、モニタ25に供給される。同様に再生された番地情報は、出力端子24b、入力端子19bを介して、メタデータの番地付与部19に供給される。メタデータ生成の担当者は、モニタ25に映出された静止画コンテンツを確認のうえ、メタデータ生成に必要な言葉をマイク16により音声入力する。それにより、カメラ11で撮影された静止画に関連する情報が、マイク16を介して音声認識部17に入力される。静止画に関連する情報は、例えば、タイトル、撮影の日時、撮影者、撮影場所（どこ）、被撮影者（だれ）、被撮影物（なに）などである。これ以降の動作は、図3の構成に関して説明したものと同様である。

(実施の形態 3)

図 6 は、本発明の実施の形態 3 におけるメタデータ作成装置の構成を示すプロック図である。本実施の形態は、一般的なデジタルデータコンテンツをメタデータ作成の対象とする例である。デジタルデータコンテンツを識別するために、コンテンツの番地あるいは番号を用いて、デジタルデータコンテンツと生成されたメタデータを関連づける構成を有する。

図 6において、3 1 はコンテンツデータベース（以下、コンテンツ DB と記す）であり、コンテンツ DB 3 1 から再生された出力は、音声入力部 3 2、ファイル処理部 3 5、および記録部 3 7 に供給される。音声入力部 3 2 の出力は、音声認識部 3 3 に供給される。音声認識部 3 3 には、辞書データベース（以下、辞書 DB と記す）3 4 のデータも供給可能である。音声認識部 3 3 からはメタデータが出力され、ファイル処理部 3 5 に入力される。ファイル処理部 3 5 では、コンテンツ DB 3 1 から供給されるタイムコード値を用いて、音声認識部 3 3 から出力されたメタデータに、所定のデータを付記した形式で file 化処理する。ファイル処理部 3 5 から出力されたメタデータファイルは記録部 3 7 に供給され、コンテンツ DB 3 1 から出力されたコンテンツとともに記録される。音声入力部 3 2 には音声入力端子 3 9 が、辞書 DB 3 4 には辞書分野選択入力端子 4 0 が、それぞれ設けられている。コンテンツ DB 3 1 からの再生出力および記録部 3 7 の再生出力は、映像モニタ 4 1 により表示可能である。

コンテンツ DB 3 1 は、例えば、VTR、ハードディスク装置、光学ディスク装置等の映像・音声信号再生装置、半導体メモリなどのメモリ手段を記録媒体とする映像・音声信号再生装置、あるいは伝送、放送により供給された映像・音声信号を 1 次記録し再生する映像・音声信号再

生装置などの、制作されたコンテンツを、コンテンツに合わせたタイムコードを発生しながら再生する機能を備えた構成を有する。

上記メタデータ作成装置の動作について、以下に説明する。コンテンツDB 3 1から再生されたタイムコード付き映像信号は、映像モニタ4 1に供給され、映出される。映像モニタ4 1に映出されたコンテンツに合わせて、作業者がマイクロホーンを用いてナレーションの音声信号を入力すると、音声入力端子3 9を介して音声信号が音声入力部3 2に入力される。

この際作業者は、映像モニタ4 1に映しだされたコンテンツ、あるいは、
10 タイムコードを確認し、シナリオ、ナレーション原稿、或いはコンテンツの内容などを基に抽出されたコンテンツ管理用キーワードを発声するようによることが好ましい。このように入力される音声信号として、前もってシナリオ等から限定されたキーワードを使用することによって、後段の音声認識部3 3での認識率を改善することができる。

15 音声入力部3 2では、コンテンツDB 1から出力されている垂直同期信号に同期したクロックで、音声入力端子3 9から入力された音声信号をデータ化する。音声入力部3 2でデータ化された音声信号データは、音声認識部3 3に入力され、また、同時に、音声認識に必要な辞書が辞書DB 3 4から供給される。辞書DB 3 4における使用する音声認識用
20 辞書は、辞書分野選択入力端子4 0から設定することができる。

例えば図7に示すように、辞書DB 3 4が各分野別に構成されていたとすると、使用する分野を辞書分野選択入力端子4 0（例えば、キー入力できるキーボード端子）から設定する。例えば料理番組の場合であれば、料理—日本料理—料理法—野菜炒め、のように、端子4 0から辞書DB 3 4の分野を設定することができる。このように辞書DB 3 4を設定することで、使用する単語、および、音声認識すべき単語を制限し、

音声認識部 3 3 の認識率を改善することができる。

また、図 6 における辞書分野選択端子 4 0 から、シナリオ、シナリオ原稿、あるいはコンテンツの内容から抽出されたキーワードを入力することができる。例えば、コンテンツが料理番組の場合であれば、図 8 に 5 示すようなレシピを端子 4 0 から入力する。コンテンツの内容を考慮すれば、レシピに記入されている単語が音声信号として入力されてくる可能性が高いので、辞書 DB 3 4 では端子 4 0 から入力されたレシピ単語の認識優先度を明示し、優先して音声認識を行うようにする。例えば、「柿」と「貝のカキ」が辞書中にあった場合、端子 4 0 から入力された 10 レシピ単語が「貝のカキ」のみの場合は、「貝のカキ」に優先順位 1 がつけられる。音声認識部 3 3 では、「かき」という音声を認識した場合、辞書 DB 3 4 に設定された単語の優先順位 1 が明記されている「貝のカキ」と認識する。

このように、辞書 DB 3 4 では、端子 4 0 から入力される分野で単語 15 を限定し、更に、シナリオを端子 4 0 から入力して単語の優先度を明示することで、音声認識部 3 3 での認識率を改善させることができる。

図 6 における音声認識部 3 3 では、辞書 DB 3 4 から供給された辞書に従って、音声入力部 3 2 から入力された音声信号データを認識し、メタデータを生成する。音声認識部 3 3 から出力されたメタデータは、ファイル処理部 3 5 に入力される。前述したように音声入力部 3 2 では、コンテンツ DB 1 から再生された垂直同期信号に同期して、音声信号をデータ化している。従って、ファイル処理部 3 5 は、音声入力部 3 2 からの同期情報と、コンテンツ DB 3 1 から供給されるタイムコード値とを用いて、例えば前述した料理番組の場合は、図 9 に示すような T E X 20 T 形式のメタデータファイルを出力する。すなわち、ファイル処理部 3 5 では、音声認識部 3 3 から出力されたメタデータに、file 開始からの 25

1秒ごとの基準時間である TM_ENT (秒) と、基準時間からのフレームオフセット数を示す TM_OFFSET と、タイムコードとを付記した形式で file 化処理する。

記録部 3 7 では、ファイル処理部 3 5 から出力されたメタデータファ

5 イルとコンテンツ DB 3 1 から出力されたコンテンツを記録する。記録部 3 7 は、HDD、メモリ、光ディスク等から構成され、コンテンツ DB 3 1 から出力されたコンテンツもファイル形式で記録する。

(実施の形態 4)

図 1 0 は、本発明の実施の形態 4 におけるメタデータ作成装置の構成

10 を示すブロック図である。本実施の形態の装置では、実施の形態 3 の構成に対して、コンテンツ情報ファイル処理部 3 6 が付加されている。コンテンツ情報ファイル処理部 3 6 では、記録部 3 7 に記録されるコンテンツの記録位置関係を示す制御ファイルを発生し、記録部 3 7 に記録する。

15 即ち、コンテンツ情報ファイル処理部 3 6 は、コンテンツ DB 3 1 から出力されたコンテンツと、記録部 3 7 から出力されるコンテンツの記録位置情報をもとに、当該コンテンツが保有している時間軸情報と、記録部 3 7 に記録したコンテンツのアドレス関係を示す情報を発生し、データ化して制御ファイルとして出力する。

20 例えば、図 1 1 に示すように、コンテンツの記録位置を示す記録メディアアドレスに対し、コンテンツの時間軸基準を示す TM_ENT #j を等時間軸間にポイントする。例えば、TM_ENT #j を 1 秒 (NTSC 信号の場合、30 フレーム) 每に記録メディアアドレスにポイントする。このようにマッピングすることで、コンテンツが 1 秒単位毎に分散記録されても、
25 TM_ENT #j から記録部 3 7 の記録アドレスを一義的に求めることができる。

さらに、図9に示したようにメタデータファイルには、ファイル開始からの1秒ごとの基準時間である TM_ENT (秒)と、基準時間からのフレームオフセット数を示す TM_OFFSET と、タイムコードと、メタデータとがTEXT形式で記録されている。従って、メタデータファイルの中で
5 メタデータ1を指定すれば、タイムコード、基準時間、及び、フレームオフセット値がわかるので、図11に示す制御ファイルから、記録部3
7での記録位置が即座にわかることになる。

なお、TM_ENT #j の等時間軸間隔は、上述のように1秒おきにポイントすることに限らず、MPEG2圧縮等で用いられているGOP単位
10 等に合わせて記述することもできる。

さらに、テレビビビジョン信号のNTSCでは垂直同期信号が
60/1.001Hz であるため、絶対時間にあわせるためにドロップフレームモードに合わせたタイムコードと、垂直同期信号 (60/1.001 Hz) にあわせたノンドロップタイムコードの2種類を使用することもできる。この場合、
15 例えば、ノンドロップタイムコードを TM_ENT #j で表し、TC_ENT #j をドロップフレーム対応タイムコードで表して使用する。

さらに、制御ファイルのデータ化は、SMIL2 等の既存言語を用いてデータ化することも可能であり、さらに、SMIL2 の機能を用いれば、関連したコンテンツ、及び、メタデータファイルのファイル名も合わせてデータ化して、制御ファイルに格納することができる。
20

さらに、図11には、記録部の記録アドレスを直接表示する構成を示したが、記録アドレスの代わりに、コンテンツファイルの冒頭からタイムコードまでのデータ容量を表示し、データ容量とファイルシステムの記録アドレスから記録部でのタイムコードの記録アドレスを計算し、検
25 出してもよい。

また、上述のように、TM_ENT #j とタイムコードの対応テーブルをメ

タデータファイルに格納する形式ではなく、TM_ENT #j とタイムコードの対応テーブルを制御ファイル中に格納しても同様の効果が得られる。

(実施の形態 5)

図 12 は、本発明の実施の形態 5 におけるメタデータ検索装置の構成 5 を示すブロック図である。本実施の形態の装置では、実施の形態 4 の構成に対して、検索部 38 が付加されている。検索部 38 では、検索したいシーンのキーワードを、音声認識してメタデータを検出するのに使用したものと同一の辞書 DB34 から選択し、設定する。

次に検索部 38 は、メタデータファイルのメタデータ項目をサーチして、キーワードと一致するタイトル名とコンテンツシーンの位置（タイムコード）の一覧を表示する。一覧表示の中からひとつの特定シーンが設定された場合は、メタデータファイルの基準時間 TM_ENT (秒) とフレームオフセット数 TM_OFFSET から、制御ファイル中の記録メディアアドレスを自動的に検出して記録部 37 に設定し、記録部 37 から当該記録メディアアドレスに記録されたコンテンツシーンをモニタ 41 に再生表示する。このように構成することで、メタデータを検出して即座に、見たいシーンを検出可能となる。

なお、コンテンツにリンクしたサムネイルファイルを準備しておけば、前述したキーワードに一致したコンテンツ名の一覧を表示する際、コンテンツの代表的サムネイル画を再生して表示することも可能である。

(実施の形態 6)

前述の実施形態 3 ~ 5 では、あらかじめ記録されているコンテンツにメタデータを付与する装置について述べたが、本実施の形態は、本発明をカメラ等、撮影時にメタデータを付与するシステム、特に、コンテンツ内容が前もって限定される風景撮りに際して、撮影位置をメタデータとして付与する装置に対して拡張した例に関する。図 13 は、本発明の

実施の形態 6 におけるメタデータ作成装置の構成を示すブロック図である。

カメラ 5 1 の撮像出力は、コンテンツ DB 5 4 に映像コンテンツとして記録される。同時に、GPS 5 2 はカメラが撮影している場所を検出し、その位置情報（経緯度数値）が音声合成部 5 3 で音声信号化されて、コンテンツ DB 5 4 の音声チャンネルに位置情報として記録される。カメラ 5 1、GPS 5 2、音声合成部 5 3、コンテンツ DB 5 4 は、記録部付きカメラ 5 0 として、一体構成することもできる。コンテンツ DB 5 4 は、音声チャンネルに記録されている音声信号の位置情報を音声認識部 5 6 に入力する。音声認識部 5 6 には、辞書 DB 5 5 から辞書データが供給される。辞書 DB 5 5 は、端子 5 9 からのキーボード入力等によって、地域名、ランドマーク等を選択、制限して、音声認識部 5 6 に出力するように構成することができる。

音声認識部 5 6 は、認識された経緯数値と辞書 DB 5 5 のデータを用いて地域名、ランドマークを検出し、ファイル処理部 5 7 に出力する。ファイル処理部 5 7 では、コンテンツ DB 5 4 から出力されたタイムコードと音声認識部 5 6 から出力された地域名、ランドマークをメタデータとして TEXT 化して、メタデータファイルを発生させる。メタデータファイルは記録部 5 8 に供給され、記録部 5 8 は、そのメタデータファイル、およびコンテンツ DB 5 4 から出力されたコンテンツデータを記録する。

このように構成することで、撮影したシーン毎に、自動的に地域名、ランドマークのメタデータを付加することができる。

上述の実施の形態では、音声認識部により認識したキーワードをタイムコードと共に、メタデータファイルでファイル化する構成を記述したが、音声認識部により認識したキーワードに加え、関連したキーワード

を追加してファイル化してもよい。例えば、音声で淀川を認識した場合は、地形、川等の一般属性キーワードも付加してファイル化する。こうすることで検索時、付加された地形、川等のキーワードも使用することができるるので検索性を向上することができる。

5 なお、本発明に係る音声認識部に、単語単位で音声認識する単語認識方式を用い、音声入力の単語数、及び、使用する認識辞書の単語数を制限することで、音声認識率を改善することができる。

また、一般的には、音声認識には何らかの誤認識が生じる可能性がある。上述の各実施形態においてキーボードを含むコンピュータ等の情報10 处理部を備え、誤認識が生じた場合に、作成されたメタデータ、あるいはタグを、キーボード操作により修正するように構成することもできる。

産業上の利用の可能性

発明のメタデータ作成装置によれば、コンテンツに関連したメタデータの作成あるいはタグ付けを行うために、音声認識を用いて音声入力によりメタデータを作成するとともに、メタデータとコンテンツの所定部位との関連付けを行うため、従来のようなキーボード入力に比べて、効率的に、メタデータの作成やタグ付けを実施することが可能となる。

請求の範囲

1. コンテンツを再生して出力するコンテンツ再生部と、
音声入力部と、
5 前記音声入力部から入力された音声信号を認識する音声認識部と、
前記音声認識部により認識された情報をメタデータに変換するメタデータ生成部と、
前記コンテンツ内の各部位を識別するための識別情報を、前記コンテンツ再生部から供給される再生されたコンテンツから取得して前記メタ
10 データに付与する識別情報付与部とを備え、
生成された前記メタデータを前記コンテンツ内の各部位と関連付ける
ように構成されたメタデータ作成装置。
2. 前記コンテンツに関連した辞書を更に備え、前記音声入力部から
15 入力された音声信号を前記音声認識部により認識する際に、前記辞書に
関連付けて認識するように構成された請求項1に記載のメタデータ作成
装置。
3. 前記音声信号を前記音声認識部により前記辞書に関連付けて単語
20 単位で認識する請求項2に記載のメタデータ作成装置。
4. キーボードを含む情報処理部を更に備え、前記キーボードからの
入力により前記情報処理部を介して前記メタデータを修正可能である請
求項1または3に記載のメタデータ作成装置。
25
5. 前記識別情報として、前記コンテンツに付与されているタイムコ

ード情報を用いる請求項 1～5 のいずれか 1 項に記載のメタデータ作成装置。

6. 前記識別情報として、前記コンテンツに付与されているコンテン
5 ツの番地、番号あるいはフレームナンバを用いる請求項 1～6 のいずれ
か 1 項に記載のメタデータ作成装置。

7. 前記コンテンツが静止画コンテンツであり、前記識別情報として、
前記静止画コンテンツの各々の番地を用いる請求項 1 に記載のメタデ
10 タ作成装置。

8. 前記コンテンツ再生部はコンテンツデータベースにより構成され、
前記音声入力部は、入力されるキーワードの音声信号を、前記コン
テンツデータベースから供給される同期信号に同期したクロックでデ
15 タ化して前記音声認識部に供給し、

前記音声認識部は、前記音声入力部でデータ化された音声信号データ
から前記キーワードを認識するように構成され、

前記メタデータ生成部は、前記識別情報として前記コンテンツに含ま
れる画像信号の時間位置を示すタイムコードを用い、前記音声認識部か
ら出力されたキーワードを前記タイムコードと結合してメタデータファ
イルを作成するファイル処理部として構成された請求項 1 に記載のメタ
データ作成装置。

9. 前記メタデータファイルとともに、前記コンテンツデータベース
25 から供給されるコンテンツをコンテンツファイルとして記録する記録部
を更に備えた請求項 8 に記載のメタデータ作成装置。

10. 前記コンテンツファイルを記録すべき記録位置と前記メタデータファイルの関係を管理する制御ファイルを発生するコンテンツ情報ファイル処理部を更に備え、

5 前記記録部に、前記コンテンツファイルおよび前記メタデータファイルとともに、前記制御ファイルを記録するように構成された請求項9に記載のメタデータ作成装置。

11. 辞書データベースを更に備え、前記音声認識部は、複数のジャンル別辞書から、前記コンテンツに適合したジャンルの辞書を選択することが可能である請求項8記載のメタデータ作成装置。

12. コンテンツに関連したキーワードを前記音声認識部に供給可能であり、前記音声認識部は、前記キーワードを優先認識するように構成された請求項11に記載のメタデータ作成装置。

13. コンテンツに関連した情報を音声入力し、入力された音声信号を音声認識装置により音声認識し、前記音声認識された情報をメタデータに変換し、前記コンテンツ内の各部位を識別するための前記コンテンツに付与された識別情報を前記メタデータに付与して、生成された前記メタデータを前記コンテンツ内の各部位と関連付けるメタデータ作成方法。

14. 前記コンテンツを再生しモニタに表示しながら、コンテンツに関連した情報を音声入力する請求項13に記載のメタデータ作成方法。

15. 前記コンテンツに関連した辞書を用い、前記入力された音声信号を前記音声認識装置により前記辞書に関連付けて認識する請求項13に記載のメタデータ作成方法。

5 16. 前記識別情報として、前記コンテンツに付与されているタイムコード情報を用いる請求項13に記載のメタデータ作成方法。

17. 前記コンテンツとして静止画コンテンツを用い、前記識別情報として、前記静止画コンテンツの各々の番地を用いる請求項13に記載
10 のメタデータ作成方法。

18. コンテンツを再生して出力するコンテンツデータベースと、
入力されるキーワードの音声信号を再生された前記コンテンツの同期
信号に同期したクロックでデータ化する音声入力部と、
15 前記音声入力部でデータ化された音声信号データからキーワードを認
識する音声認識部と、前記音声認識部から出力されたキーワードを前記
コンテンツに含まれる画像信号の時間位置を示すタイムコードと結合さ
せてメタデータファイルを作成するファイル処理部と、

20 コンテンツファイルの記録位置と前記メタデータファイルの関係を管
理する制御ファイルを発生するコンテンツ情報ファイル処理部と、
前記コンテンツファイル、前記メタデータファイル、および前記制御
ファイルを記録する記録部と、入力される検索用キーワードが含まれた
前記メタデータファイルを特定し、前記制御ファイルを参照して、前記
コンテンツファイルの前記キーワードに対応する記録位置を抽出する検
索部とを備え、
25 前記コンテンツファイルの記録位置は、前記記録部における記録位置

であるメタデータ検索装置。

19. 前記コンテンツ情報ファイル処理部から出力される制御ファイルは、前記コンテンツの記録時間に合わせた前記記録部における前記コンテンツの記録位置を明記するテーブルとし、前記タイムコードから前記コンテンツの記録位置を検索できるようにした請求項18記載のメタデータ検索装置。

20. 辞書データベースと、コンテンツに関連したキーワードを前記音声認識部に供給するキーワード供給部とを更に備え、前記音声認識部は、複数のジャンル別辞書から、前記コンテンツに適合したジャンルの辞書を選択することが可能であるとともに、前記キーワードを優先認識するように構成された請求項18記載のメタデータ検索装置。

21. 辞書データベースを更に備え、前記音声認識部は、複数のジャンル別辞書から、前記コンテンツに適合したジャンルの辞書を選択することが可能であるとともに、

前記検索部は、前記音声認識部で用いた共通辞書から選定されたキーワードにより検索を行うように構成された請求項18記載のメタデータ検索装置。

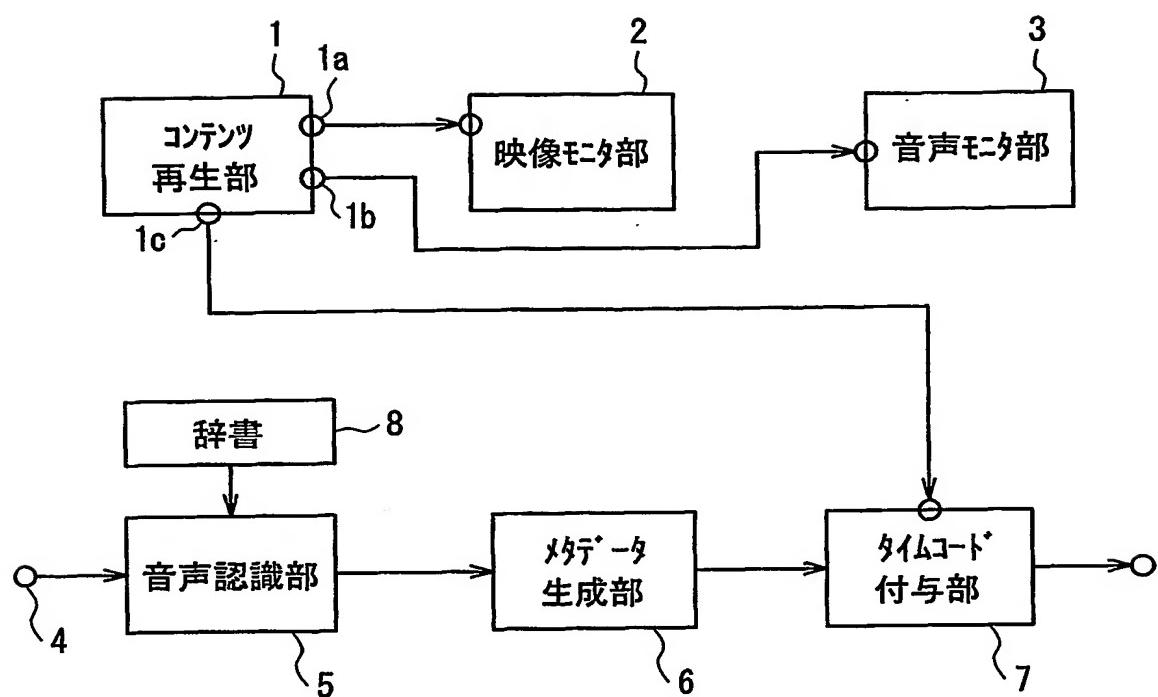


FIG. 1

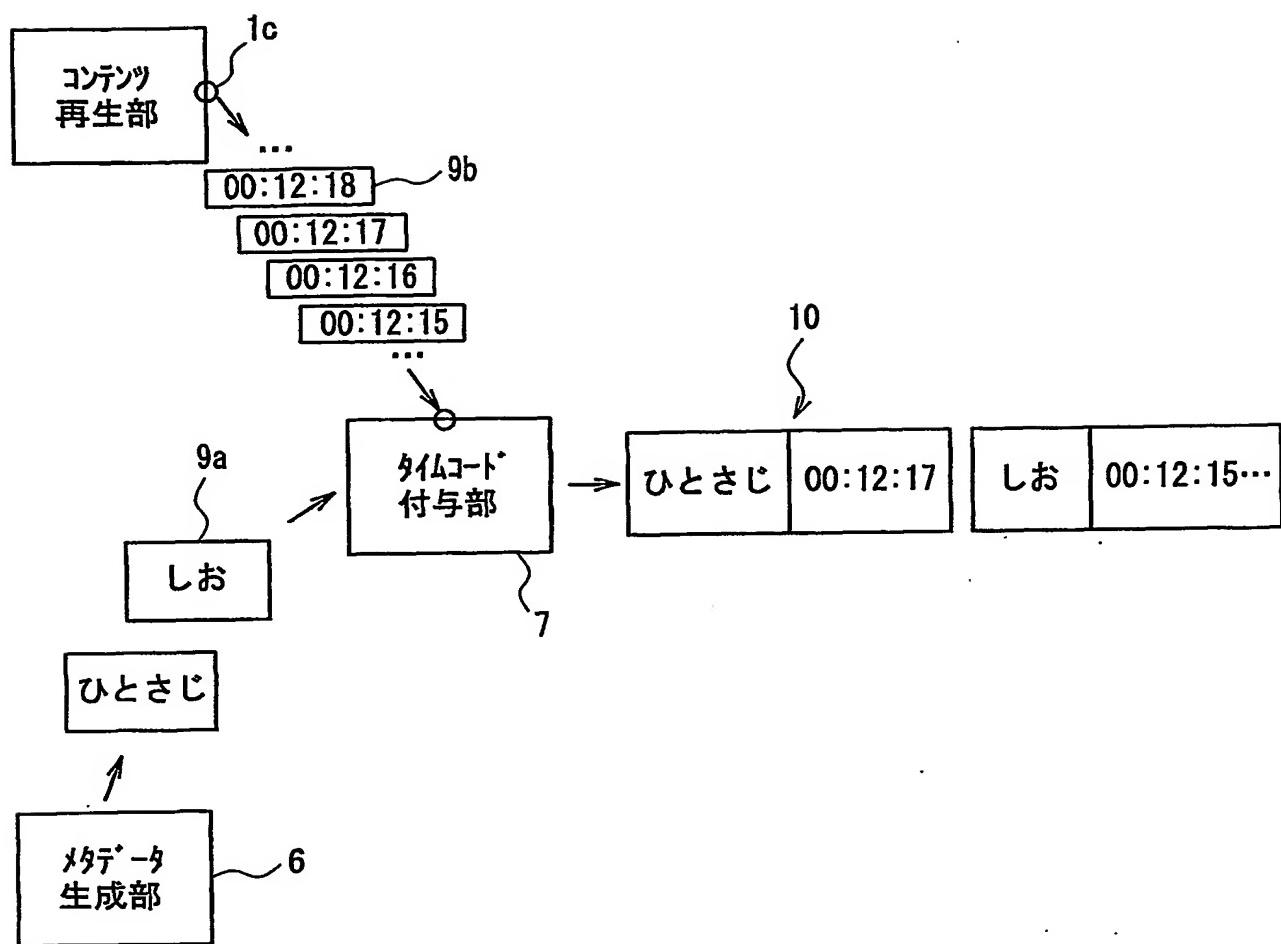


FIG. 2

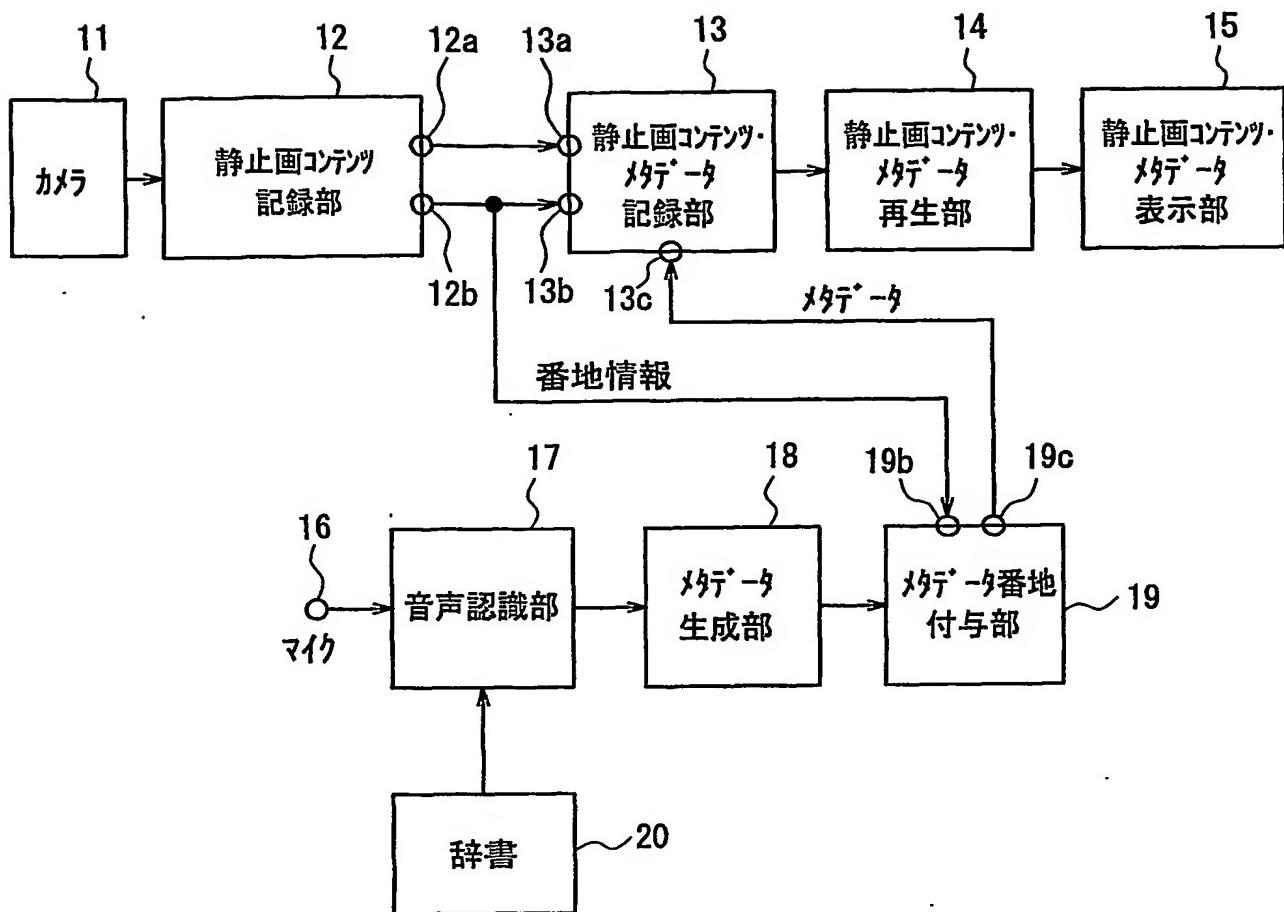


FIG. 3

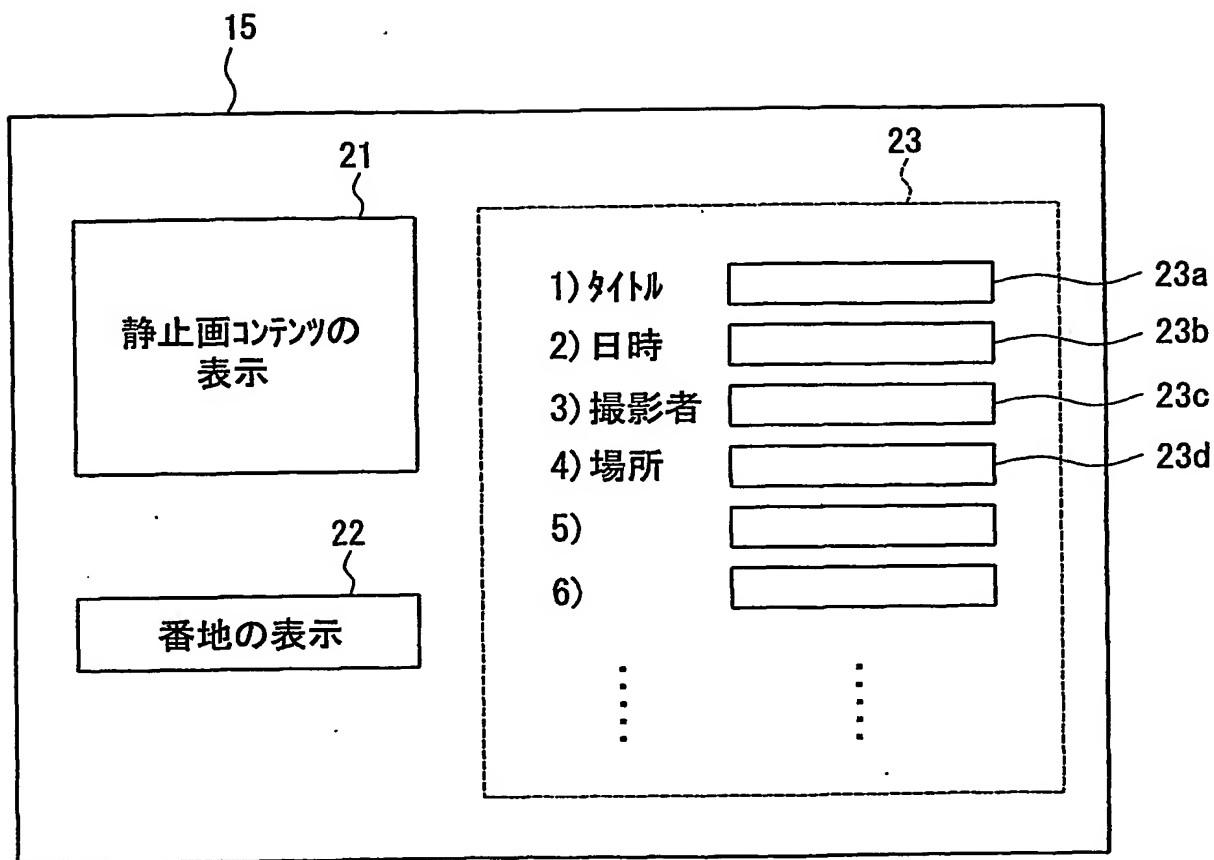


FIG. 4

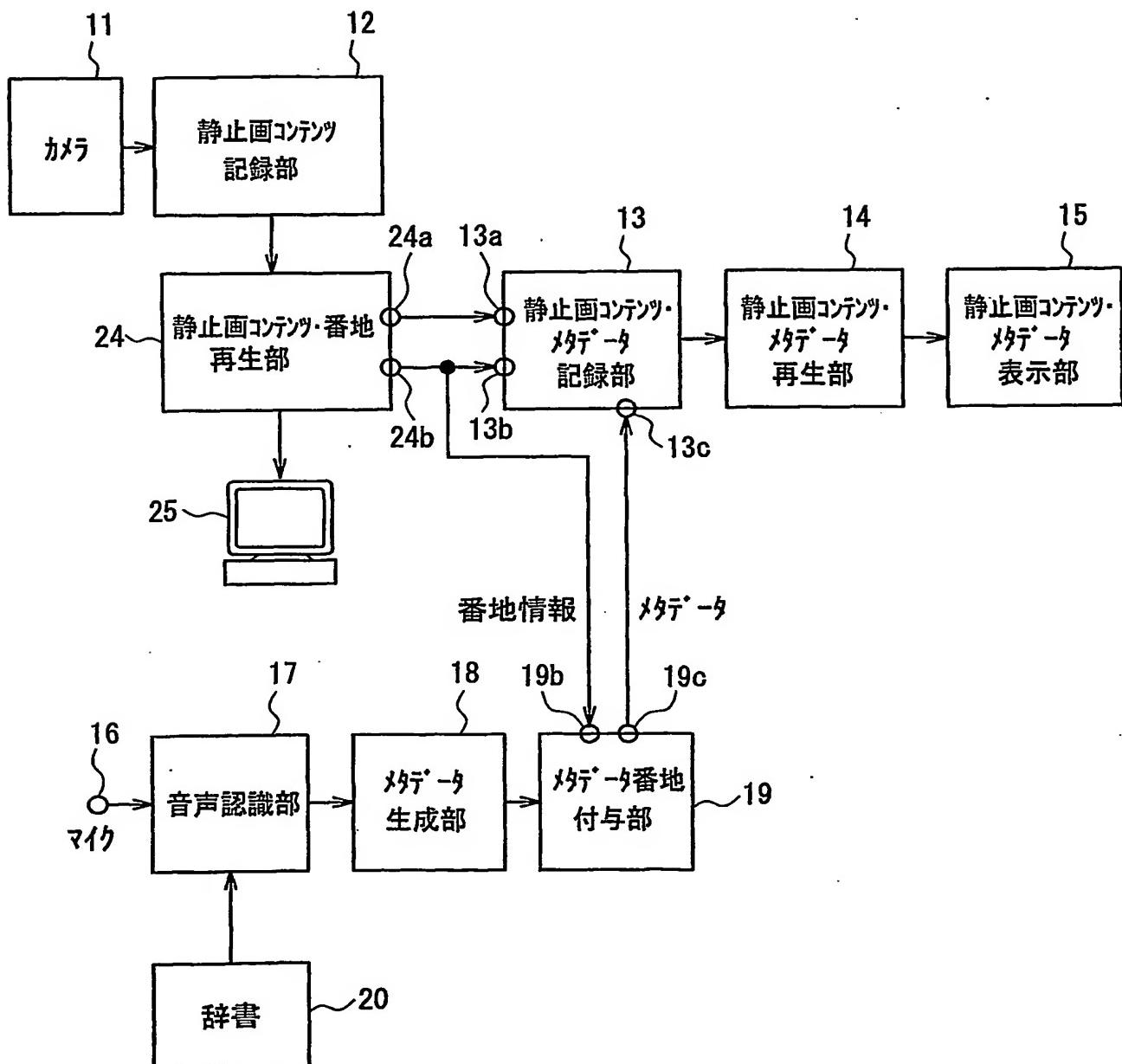


FIG. 5

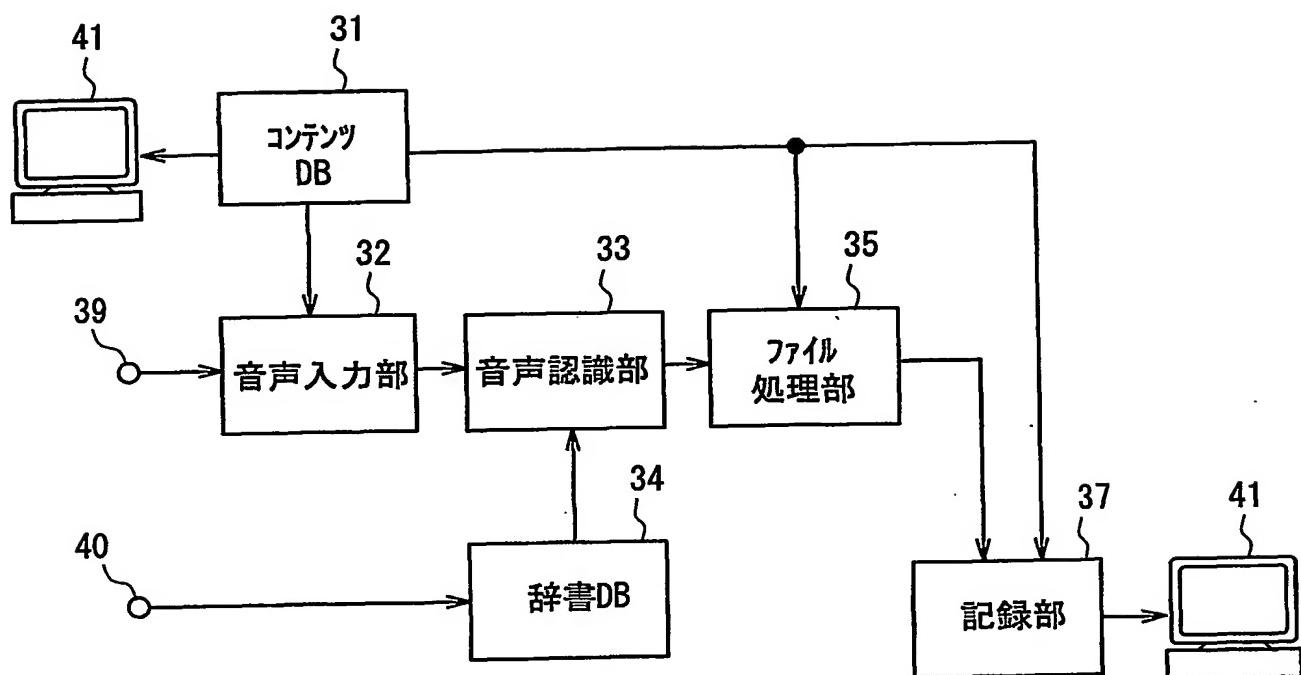


FIG. 6

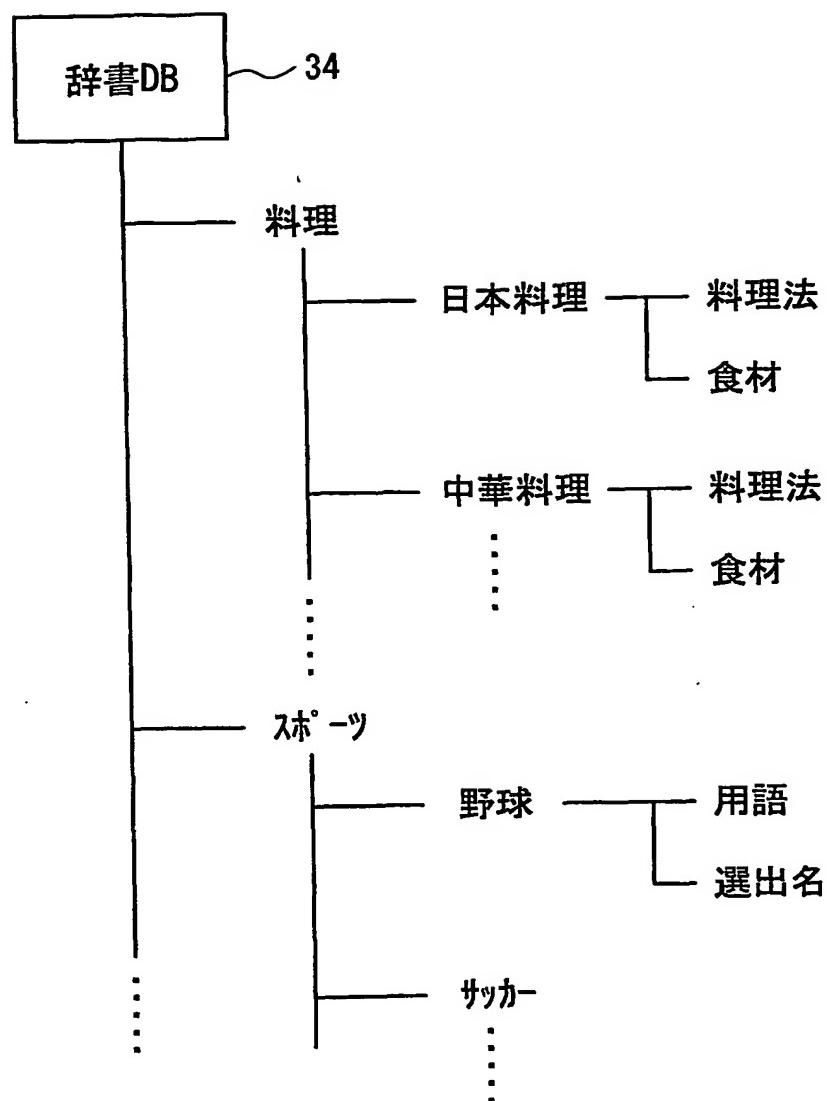


FIG. 7

レシピ
＜メニュー＞ アサリのトマト煮 白ワイン風味

＜材料＞ 4人分

材料	分量
あさり	50個
カキ	10個
玉ねぎ(小)	1個分
トマト(湯むきし小さく刻んだもの)	1個分
にんにく(みじん切り)	2枝
白ワイン	1カップ
タイム	少々
パセリ(みじん切り)	1/2カップ
オリーブ油	1/4カップ

FIG. 8

Title:あさりのトマト煮

TM_ENT#j	TM_OFFSET*1	TC	メタデータ
930	+8	00 15 30 08	玉ねぎ 1個
990	+5	00 16 30 05	にんにく 2枝
1230	+10	00 20 30 10	あさり、カキ 炒める

*1:オフセットフレーム数で表記

FIG. 9

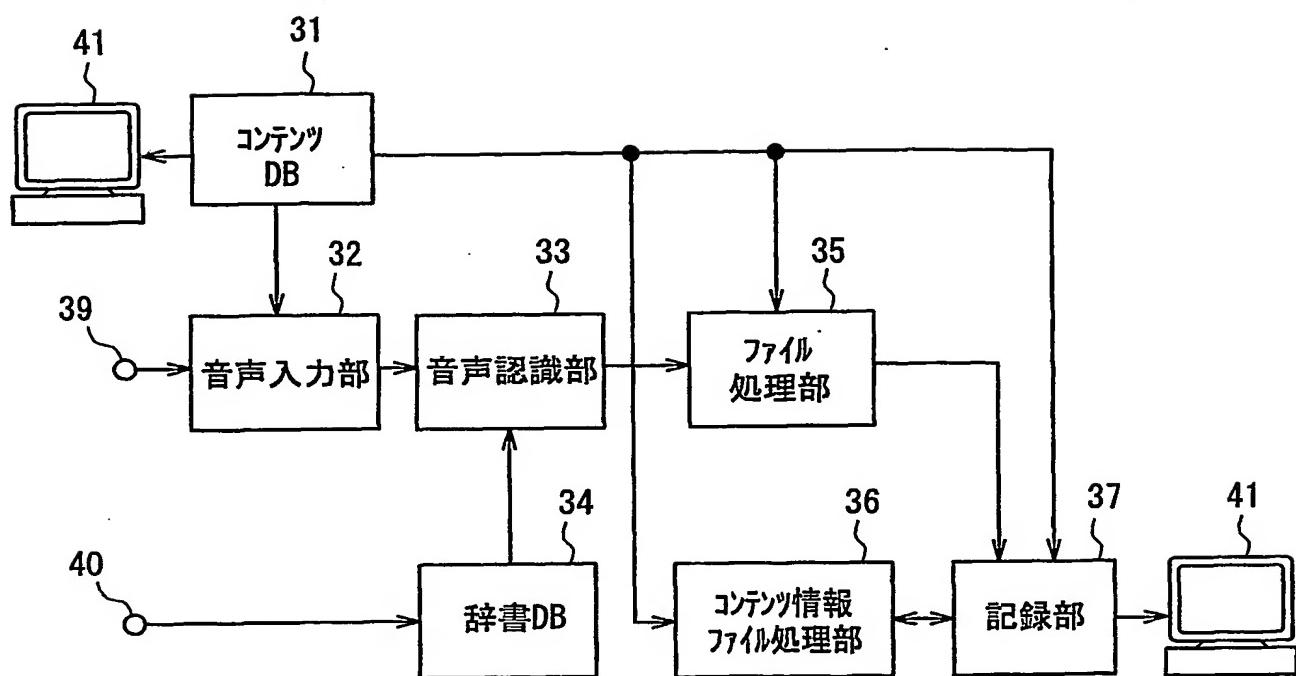


FIG. 10

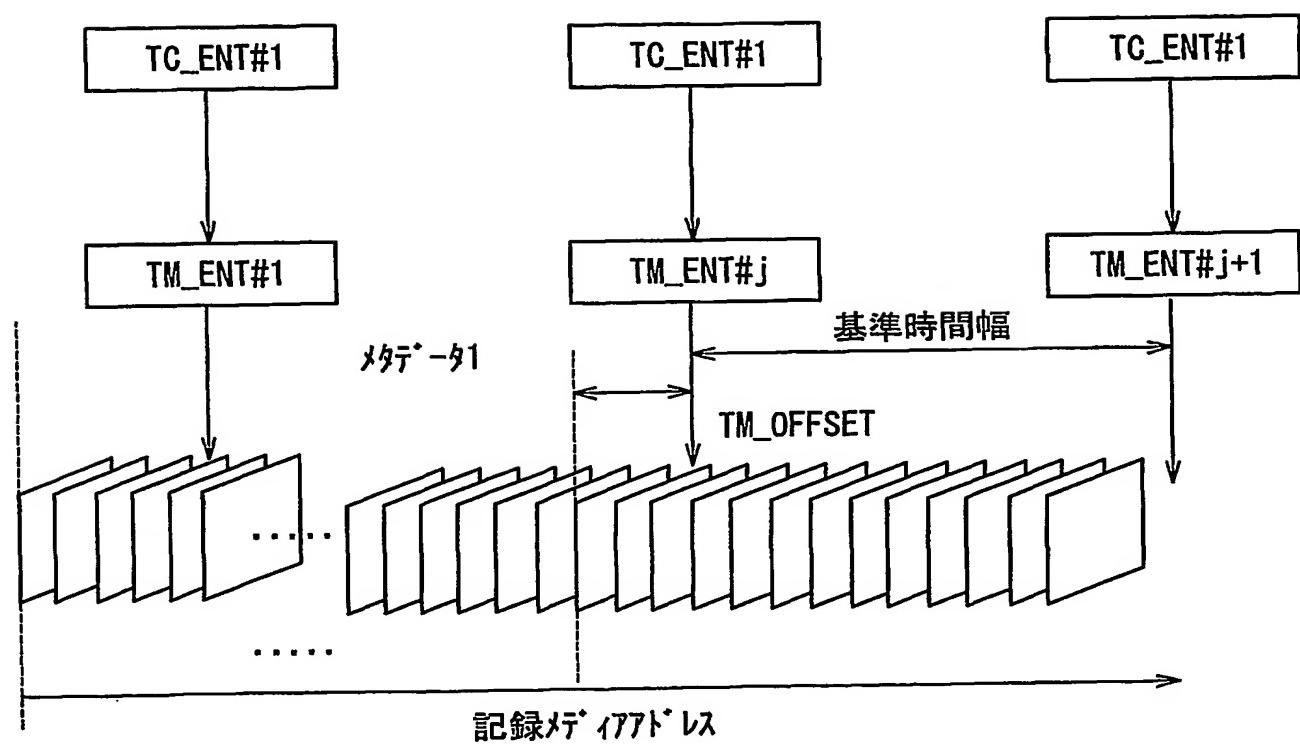


FIG. 11

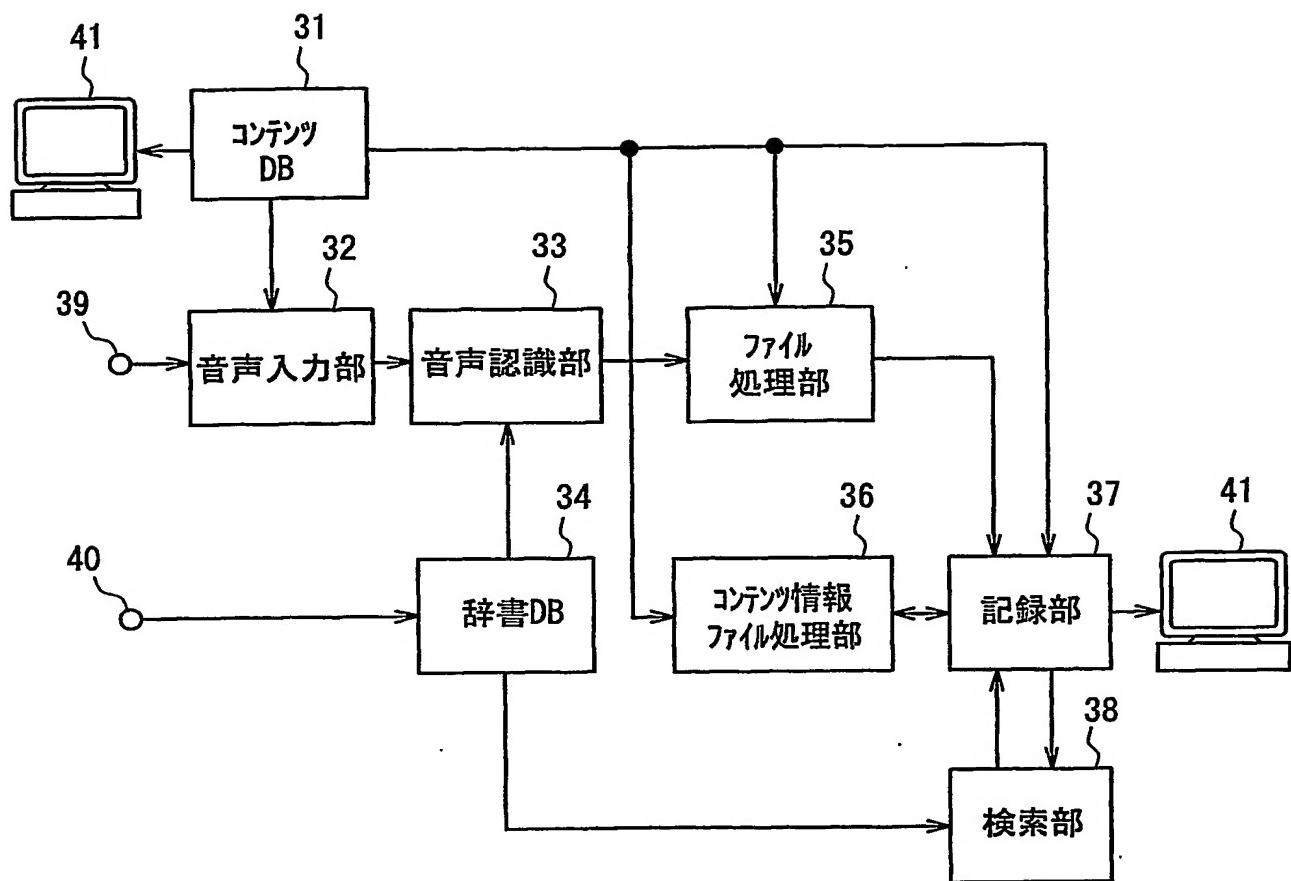


FIG. 12

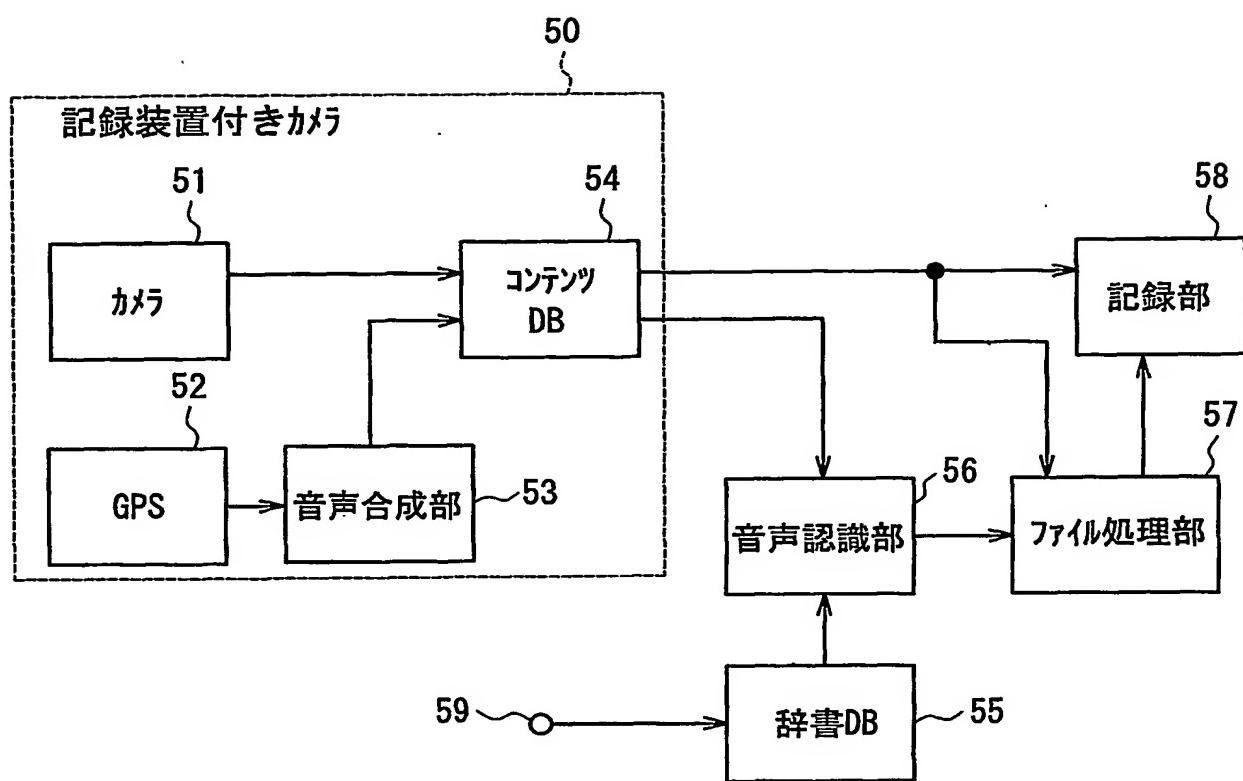


FIG. 13

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP03/07908

A. CLASSIFICATION OF SUBJECT MATTER
Int.Cl⁷ H04N5/76

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl⁷ H04N5/76-5/956

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2003
Kokai Jitsuyo Shinan Koho	1971-2003	Toroku Jitsuyo Shinan Koho	1994-2003

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X A	JP 2003-111009 A (Fuji Photo Film Co., Ltd.), 11 April, 2003 (11.04.03), Full text; Figs. 1 to 7 (Family: none)	1-10, 13-19 11, 12, 20, 21
P, X A	JP 2003-32625 A (Canon Inc.), 31 January, 2003 (31.01.03), Full text; Figs. 1 to 13 (Family: none)	13, 15-17 1-12, 14, 18-21
P, X A	JP 2002-374494 A (Fuji Electric Co., Ltd.), 26 December, 2002 (26.12.02), Full text; Figs. 1 to 5 (Family: none)	1-10, 13-19 11, 12, 20, 21

 Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:	
"A"	document defining the general state of the art which is not considered to be of particular relevance
"E"	earlier document but published on or after the international filing date
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O"	document referring to an oral disclosure, use, exhibition or other means
"P"	document published prior to the international filing date but later than the priority date claimed
"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&"	document member of the same patent family

Date of the actual completion of the international search 27 October, 2003 (27.10.03)	Date of mailing of the international search report 11 November, 2003 (11.11.03)
--	--

Name and mailing address of the ISA/ Japanese Patent Office	Authorized officer
Facsimile No.	Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP03/07908

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X A	JP 2002-171481 A (Ricoh Co., Ltd.), 14 June, 2002 (14.06.02), Full text; Figs. 1 to 3 (Family: none)	1-10, 13-19 11, 12, 20, 21
X A	JP 2000-306365 A (Sony Corp.), 02 November, 2000 (02.11.00), Full text; Figs. 1 to 20 (Family: none)	1-10, 13-19 11, 12, 20, 21
X A	JP 2000-78530 A (NEC Corp.), 14 March, 2000 (14.03.00), Full text; Figs. 1 to 4 (Family: none)	13, 15-17 1-12, 14, 18-21
X A	JP 09-149365 A (Ricoh Co., Ltd.), 06 June, 1997 (06.06.97), Par. Nos. [0060] to [0065]; Figs. 1 to 13 & US 6111605 A1 & US 6380975 B1 & DE 19645716 A	1-10, 13-19 11, 12, 20, 21
X A	JP 09-130736 A (Sony Corp.), 16 May, 1997 (16.05.97), Full text; Figs. 1 to 13 (Family: none)	13, 15-17 1-12, 14, 18-21
P,A	JP 2003-18505 A (Toshiba Corp.), 17 January, 2003 (17.01.03), Full text; Figs. 1 to 17 (Family: none)	1-21
P,A	JP 2002-207753 A (Teijin Seiki Co., Ltd.), 26 July, 2002 (26.07.02), Full text; Figs. 1 to 6 (Family: none)	1-21
A	JP 07-184160 A (Canon Inc.), 21 July, 1995 (21.07.95), Full text; Figs. 1 to 13	1-21

A. 発明の属する分野の分類（国際特許分類（IPC））

Int. Cl' H04N 5/76

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（IPC））

Int. Cl' H04N 5/76-5/956

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922～1996年
日本国公開実用新案公報	1971～2003年
日本国実用新案登録公報	1996～2003年
日本国登録実用新案公報	1994～2003年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
P X A	JP 2003-111009 A (富士写真フィルム株式会社) 2003.04.11 全文、図1-7 (ファミリーなし)	1-10, 13-19 11, 12, 20, 21
P X A	JP 2003-32625 A (キャノン株式会社) 2003.01.31 全文、図1-13 (ファミリーなし)	13, 15-17 1-12, 14, 18-21
P X A	JP 2002-374494 A (富士電機株式会社) 2002.12.26 全文、図1-5 (ファミリーなし)	1-10, 13-19 11, 12, 20, 21

 C欄の続きにも文献が列挙されている。 パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

- 「A」特に関連のある文献ではなく、一般的技術水準を示すもの
- 「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
- 「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
- 「O」口頭による開示、使用、展示等に言及する文献
- 「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

- 「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
- 「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
- 「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
- 「&」同一パテントファミリー文献

国際調査を完了した日

27.10.03

国際調査報告の発送日

11.11.03

国際調査機関の名称及びあて先

日本国特許庁 (ISA/JP)

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官（権限のある職員）

松元 伸次



5 C

9563

電話番号 03-3581-1101 内線 3540

C(続き)、 引用文献の カテゴリー*	関連すると認められる文献 引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
X	JP 2002-171481 A (株式会社リコー) 2002.06.14	1-10, 13-19
A	全文、図1-3 (ファミリーなし)	11, 12, 20, 21
X	JP 2000-306365 A (ソニー株式会社) 2000.11.02	1-10, 13-19
A	全文、図1-20 (ファミリーなし)	11, 12, 20, 21
X	JP 2000-78530 A (日本電気株式会社) 2000.03.14	13, 15-17
A	全文、図1-4 (ファミリーなし)	1-12, 14, 18-21
X	JP 09-149365 A (株式会社リコー) 1997.06.06 【0060】～【0065】、図1-13 & US 6111605 A1, & US 6380975 B1, & DE 19645716 A	1-10, 13-19 11, 12, 20, 21
X	JP 09-130736 A (ソニー株式会社) 1997.05.16 全文、図1-13 (ファミリーなし)	13, 15-17 1-12, 14, 18-21
P A	JP 2003-18505 A (株式会社東芝) 2003.01.17 全文、図1-17 (ファミリーなし)	1-21
P A	JP 2002-207753 A (帝人製機株式会社) 2002.07.26 全文、図1-6 (ファミリーなし)	1-21
A	JP 07-184160 A (キャノン株式会社) 1995.07.21 全文、図1-13	1-21